An Effective TU Size Decision Method for Fast HEVC Encoders

Chou-Chen Wang¹, Yen-Czu Liao¹, Jing-Wein Wang², Chi-Wei Tung¹

¹Department of Electronic Engineering, I-Shou University, Kaohsiung, Taiwan
²Institute of Photonics and Communications, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan

E-mail: chchwang@isu.edu.tw

Abstract—In order to speed up the encoding process of HEVC, there have been many fast encoding methods proposed to reduce the number of CUs and PUs. Besides, the early TU decision algorithm (ETDA) is another method selected to reduce the encoding complexity of TU. Recently, Chio et al. proposed a new ETDA by determining the number of nonzero DCT coefficients (NNZ) of RQT (called NNZ-ETDA) to accelerate the encoding process of TU module [6]. However, the NNZ-ETDA can’t effectively reduce the computational load for sequences with active motion or rich texture. Therefore, in order to further improve the performance of NNZ-ETDA, we propose an adaptive RQT depth for NNZ-ETDA (called ARD-NNZ-ETDA) by exploiting the characteristics of high temporal-spatial correlation exists in nature video sequences. An adaptive depth of RQT is employed to the NNZ-ETDA to further reduce the computational load of TU. Simulation results show that the proposed method can achieve time improving ratio (TIR) about 61.26%~81.48% when compared to HEVC (HM 8.1) with insignificant loss of image quality. Compared with the NNZ-ETDA, the proposed method can further achieve an average TIR about 8.29%~17.92%.

Keywords—HEVC, DCT, RQT, temporal-spatial correlation

I. INTRODUCTION

With the rapid development of electronic technology, the panels of 4K×2K (or 8K×4K) high-resolution will become the main specification of large-size digital TV in future. However, the current H.264 video coding standard can’t support the video applications of high definition (HD) and ultrahigh definition (UHD) resolution. Therefore, the ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Pictures Expert Group (MPEG) through their Joint Collaborative Team on Video Coding (JCT-VC) has been developed a newest high efficiency video coding (HEVC) for video compression standard to satisfy the UHD requirement in 2010, and the first version of HEVC was finalized by JCT-VC in 2013 [1-2].

The compression efficiency of HEVC is two times better when compared to that of H.264/AVC. This is because it adopts new techniques including hierarchical quadtree structure of coding unit (CU), prediction unit (PU) and transformation unit (TU). The CU size ranges from largest CU (LCU: 64×64) to the smallest CU (SCU: 8×8) pixels and TUs vary from 32×32 to 4×4 pixels. The relationship between the CU, PU and TU is shown in Fig.1. HEVC encoder enables 7 different inter-partition modes including SKIP, inter 2N×2N, inter 2N×N, inter N×2N, intra 2N×2N and intra N×N are used for the PUs as shown in Fig.2. The rate distortion (RD) cost under all partition modes and all CU sizes has to be calculated by performing the PUs and TUs to select the optimal CU size and partition mode. However, this “try all and select the best” method will result in the high computational complexity and limit the use of HEVC encoders in real-time applications.

In order to speed up the encoding process of HEVC, there have been many fast encoding methods proposed to reduce the
encoding complexity [6]. In HEVC, every possible CU size is number of CUs and PUs [3-5]. Besides, the early TU decision algorithm (ETDA) is another method selected to reduce the tested in order to estimate the coding performance of each CU, the coding performance is determined by the CU size compared with the corresponding PUs and TUs. Among them, TU is the basic unit used for the transform and quantisation processes. The TU process recursively partitions the structure of a given prediction block (from PU) into transform blocks that are represented by the residual quadtree (RQT). Recently, Chio et al. proposed a new ETDA by determining the number of nonzero DCT coefficients (NNZ) of RQT (NNZ-ETDA) to accelerate the encoding process of TU module [6]. However, the NNZ-ETDA can’t effectively reduce the computational load for sequences with active motion or rich texture. Therefore, in order to further improve the performance of NNZ-ETDA, we propose an adaptive RQT depth for NNZ-ETDA (ARD-NNZ-ETDA) by exploiting the characteristics of high temporal-spatial correlation exists in nature video sequences. Firstly, we analyse and calculate the temporal and spatial correlation values from the temporal (co-located) and spatial (left, upper and left upper) neighbouring blocks of the current TU. And then, the dynamic depth level range of RQT of the current TU is predicted by using the correlation weight and maximum depth levels of its neighbouring blocks. Finally, we combine the proposed adaptive depth of RQT and the NNZ-ETDA to further reduce the computational load of TU.

II. OVERVIEW OF NNZ-EDTA

Although the coding efficiency in HEVC can be improved by using various transform block sizes, the computational complexity of RD cost function increased dramatically. This is because that results from testing all the TUs. Figure 3 shows the example of optimal RQT structure given CU size of 32×32, the RD cost evaluation is performed a number of times within each RQT structure: once for the 32×32 TU, four times for the 16×16 TU, and 16 times for the 8×8 TU. In order to further speed up the encoding process of HEVC, Chio et al. proposed an early TU decision method for fast video encoding by pruning the TUs at an early stage based on the number of nonzero DCT coefficients (NNZ-ETDA).

Chio et al. find when a reasonable cost is found at an early stage, the encoder is not able to skip the remaining RQT processes by using sub-tree computations in the test model of HEVC (HM). On the other hand, the \( P_r(X_k|NZ_N = k) \) is used to calculate the conditional probability of the selection of a root node \((X_0)\) given NNZ \((NZ_N = k)\) where \(N\) is the TU size of the root node. They further find that a strong correlation exists in between the determined TU size and NNZ. Therefore, they proposed NNZ-EDTA to early determine the TU size and to stop the RD cost evaluation. NNZ is selected as a threshold to stop further RD cost evaluation based on balancing the computational complexity reduction and the compression efficiency loss. However, from the experimental results, we find that the NNZ-ETDA can’t effectively reduce the computational load for sequences with fast motion or rich texture. Table I and Table II show the conditional probability distributions for sequences. For sequences contain a large area of low activities, more than 70% blocks select the depth level “1” as the optimal levels. These results show that small depth level distributions for sequences. For sequences contain a large area of low activities, more than 70% blocks select the depth level “1” as the optimal levels. These results show that small depth levels are always selected at TUs in the homogeneous region, and large depth levels are selected at TUs with active motion or rich texture. The depth range should be adaptively determined based on the residual block property.

### Table I

| OP | Conditional probability \( P_r(X_k|NZ_N = k) \) (%) |
|----|--------------------------------------------------|
| N = 32 | \( k=0 \) | \( k=1 \) | \( k=2 \) | \( k=3 \) |
| 22 | 98.5 | 62.2 | 49.1 | 40.1 |
| 27 | 94.0 | 65.0 | 58.2 | 44.8 |
| 32 | 96.5 | 81.0 | 68.9 | 54.7 |
| 37 | 97.8 | 80.8 | 65.9 | 57.6 |
| Ave. | 96.7 | 72.3 | 60.5 | 49.3 |

### Table II

| OP | Conditional probability \( P_r(X_k|NZ_N = k) \) (%) |
|----|--------------------------------------------------|
| N = 32 | \( k=0 \) | \( k=1 \) | \( k=2 \) | \( k=3 \) | \( k=0 \) | \( k=1 \) | \( k=2 \) | \( k=3 \) | \( k=0 \) | \( k=1 \) | \( k=2 \) | \( k=3 \) | \( k=0 \) | \( k=1 \) | \( k=2 \) | \( k=3 \) |
| 22 | 81.5 | 62.5 | 53.5 | 40.4 | 84.3 | 80.7 | 71.2 | 69.5 | 92.1 | 90.1 | 83.6 | 81.0 |
| 27 | 84.5 | 63.3 | 49.8 | 38.6 | 88.5 | 81.5 | 71.0 | 65.2 | 95.1 | 92.0 | 83.8 | 80.3 |
| 32 | 85.8 | 77.1 | 53.9 | 40.6 | 91.5 | 86.6 | 75.8 | 70.1 | 97.2 | 94.1 | 86.4 | 81.6 |
| 37 | 87.7 | 73.7 | 54.0 | 41.3 | 94.1 | 88.9 | 78.7 | 72.9 | 98.6 | 94.8 | 87.4 | 81.2 |
| Ave. | 85.1 | 69.2 | 52.8 | 40.2 | 89.6 | 84.4 | 74.2 | 68.9 | 95.8 | 92.8 | 85.3 | 81.0 |

![Fig.3 Example of RQT for dividing given coding tree block.](image)

![Fig.3 Temporal and spatial correlations of TUs.](image)
Natural video sequences have strong spatial and temporal correlations, especially in the homogeneous regions. The optimal RQT depth level of a certain TU is the same or very close to the depth level of its spatially adjacent blocks due to the high correlation between adjacent blocks [7]. Therefore, we firstly analyse and calculate the temporal and spatial correlation values of maximum RQT depth from the temporal (co-located: Col) and spatial (left: L, upper: U and left upper: L-U) neighbouring blocks of the current TU as shown in Fig. 4. Table III shows the probability of the same maximum RQT depth among the temporal (Col) and spatial (L, U and L-U) neighbouring blocks of the current TU.

To speed up RQT pruning, we make use of spatial and temporal correlations to analyse region properties and skip unnecessary TU sizes. Specifically, the optimal RQT depth level of a block is predicted using spatial neighbouring blocks and the co-located block at the previously coded frame as follows:

\[
\text{Depth}_{\text{pred}} = \sum_{i=0}^{N} a_i \cdot d_i
\]

where \(N\) is the number of blocks equal to 4, \(d_i\) is the value of depth level and \(a_i\) is the weight determined based on correlations between the current block and its neighbouring blocks. The four weights are normalized to have \(\sum_{i=0}^{N} a_i = 1\). From Table III, we observe that the co-located block in the previously coded frame and spatial neighbouring blocks have almost the same correlation of the maximum RQT depth for the current TU after normalization. Therefore, the weights for four neighbouring blocks are set to 0.25.

According to the predicted value of the optimal RQT depth, each block is divided into five types as follows:

1. If \(\text{Depth}_{\text{pred}} < 1\), its optimal RQT depth is chosen to “1” and the TU size is 32×32. The dynamic depth range (DR) of current TU is classified as Type 0.
2. If \(1 \leq \text{Depth}_{\text{pred}} \leq 1.5\), its optimal RQT depth is chosen to “2” and the TU size is 32×32–16×16. The dynamic DR of current TU is classified as Type 1.
3. If \(1.5 \leq \text{Depth}_{\text{pred}} \leq 2.5\), its optimal RQT depth is chosen to “3” and the TU size is 32×32–8×8. The dynamic DR of current TU is classified as Type 2.
4. If \(2.5 \leq \text{Depth}_{\text{pred}} \leq 3.5\), its optimal RQT depth is chosen to “4” and the TU size is 16×16–4×4. The dynamic DR of current TU is classified as Type 3.
5. If \(\text{Depth}_{\text{pred}} > 3.5\), its optimal RQT depth is chosen to “4” and the TU size is 8×8–4×4. The dynamic DR of current TU is classified as Type 4.

(2) If \(1 \leq \text{Depth}_{\text{pred}} \leq 1.5\), its optimal RQT depth is chosen to “2” and the TU size is 32×32–16×16. The dynamic DR of current TU is classified as Type 1.

(3) If \(1.5 \leq \text{Depth}_{\text{pred}} \leq 2.5\), its optimal RQT depth is chosen to “3” and the TU size is 32×32–8×8. The dynamic DR of current TU is classified as Type 2.

(4) If \(2.5 \leq \text{Depth}_{\text{pred}} \leq 3.5\), its optimal RQT depth is chosen to “4” and the TU size is 16×16–4×4. The dynamic DR of current TU is classified as Type 3.

(5) If \(\text{Depth}_{\text{pred}} > 3.5\), its optimal RQT depth is chosen to “4” and the TU size is 8×8–4×4. The dynamic DR of current TU is classified as Type 4.

Based on the above analysis, the candidate depth levels that will be tested using RDO for each TU are summarized in Table VI. Therefore, the dynamic depth level range of RQT of the current TU can be predicted by using the correlation weight and maximum depth levels of its neighbouring blocks. Finally, we combine the proposed adaptive depth of RQT and the NNZ-ETDA to further reduce the computational load of TU. In other words, we propose an adaptive RQT depth to set the candidate depth levels before performing the NNZ-ETDA.

### IV. EXPERIMENTAL RESULTS

The coding performance is evaluated by the comparisons of \(\Delta\text{Bitrate}, \Delta\text{PSNR} \) and time improving ratio (TIR) between the HEVC test model (HM 8.1) [8–9], NNZ-ETDA (threshold \(k=3\)) and the proposed method. \(\Delta\text{Bitrate}, \Delta\text{PSNR} \) and TIR are respectively defined as follows:

\[
\Delta\text{Bitrate} = \frac{\text{Bitrate}_{\text{method}} - \text{Bitrate}_{\text{HM 8.1}}}{\text{Bitrate}_{\text{HM 8.1}}} \times 100 \%
\]

\[
\Delta\text{PSNR} = \text{PSNR}_{\text{method}} - \text{PSNR}_{\text{HM 8.1}}
\]

\[
\text{TIR} = \frac{\text{Bitrate}_{\text{HM 8.1}} - \text{Bitrate}_{\text{method}}}{\text{Bitrate}_{\text{HM 8.1}}} \times 100 \%
\]
V. CONCLUSIONS

We propose an adaptive RQT depth for NNZ-ETDA by exploiting the characteristics of high temporal-spatial correlation exists in nature video sequences. An adaptive depth of RQT is employed to the NNZ-ETDA to further reduce the computational load of TU. The proposed method can achieve time improving ratio about 61.26%–81.48% as compared to HM 8.1 encoder with insignificant loss of image quality. When compared with the NNZ-ETDA, the proposed method can further achieve time improving ratio about 8.29%–17.92%. In addition, the proposed method also can be equally implemented with or be considered in design of a fast HEVC encoder.

ACKNOWLEDGMENTS

This work was supported by the I-Shou University, Taiwan, R.O.C., under Contract ISU 103-01-05A.

REFERENCES